www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

HPC CHALLENGES IN THE OIL INDUSTRY, PART 1: GEOPHYSICAL IMAGING TOOLS

J. de la Puente, BSC-CASE

Top500.org - November 2015

13	NASA/Ames Research Center/NAS United States	**Pleiades** - SGI ICE X, Intel Xeon E5-2670/E5-2680v2/E5-2680v3 2.6/2.8/2.5 GHz, Infiniband FDR SGI	185,344	4,089.4	4,970.9	3,380
14	Petroleum Geo-Services United States	**Abel** - Cray XC30, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect Cray Inc.	145,920	4,042.5	5,369.9	1,800
15	Government United States	Cray CS-Storm, Intel Xeon E5-2660v2 10C 2.2GHz, Infiniband FDR, Nvidia K40 2.3GHz, Infiniband FDR SGI	72,800	3,577.0	6,131.8	1,499
19	Exploration & Production - Eni S.p.A. Italy	**HPC2** - iDataPlex DX360M4, Intel Xeon E5-2680v2 10C 2.8GHz, Infiniband FDR, NVIDIA K20x IBM	72,000	3,188.0	4,605.0	1,227
20	Government United States	Cray XC30, Intel Xeon E5-2697v2 12C 2.7GHz, Aries interconnect Fujitsu	225,984	3,143.5	4,881.3	
32	Saudi Aramco Saudi Arabia	**Makman-2** - Dell PowerEdge R630, Xeon E5-2680v3 12C 2.5GHz, Infiniband QDR Dell	76,032	2,249.7	3,041.3	1,134
33	Total Exploration Production France	**Pangea** - SGI ICE X, Xeon E5-2670 8C 2.600GHz, Infiniband FDR SGI	110,400	2,098.1	2,296.3	2,118
34	LvLiang Cloud Computing Center China	**Tianhe-2 LvLiang Solution** - Tianhe-2 LvLiang, Intel Xeon E5-2692v2 12C 2.2GHz,	174,720	2,071.4	3,074.5	997

Top500.org - November 2015

13	NASA/Ames Research Center/NAS United States	**Pleiades** - SGI ICE X, Intel Xeon E5-2670/E5-2680v2/E5-2680v3 2.6/2.8/2.5 GHz, Infiniband FDR SGI	185,344	4,089.4	4,970.9	3,380
14	Petroleum Geo-Services United States	**Abel** - Cray XC30, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect Cray Inc.	145,920	4,042.5	5,369.9	1,800
15	Government United States	Cray CS-Storm, Intel Xeon E5-2660v2 10C 2.2GHz, Infiniband FDR, Nvidia K40 2.3GHz, Infiniband FDR SGI	72,800	3,577.0	6,131.8	1,499
19	Exploration & Production - Eni S.p.A. Italy	**HPC2** - iDataPlex DX360M4, Intel Xeon E5-2680v2 10C 2.8GHz, Infiniband FDR, NVIDIA K20x IBM	72,000	3,188.0	4,605.0	1,227
20	Government United States	Cray XC30, Intel Xeon E5-2697v2 12C 2.7GHz, Aries interconnect Fujitsu	225,984	3,143.5	4,881.3	
32	Saudi Aramco Saudi Arabia	**Makman-2** - Dell PowerEdge R630, Xeon E5-2680v3 12C 2.5GHz, Infiniband QDR Dell	76,032	2,249.7	3,041.3	1,134
33	Total Exploration Production France	**Pangea** - SGI ICE X, Xeon E5-2670 8C 2.600GHz, Infiniband FDR SGI	110,400	2,098.1	2,296.3	2,118
34	LvLiang Cloud Computing Center China	**Tianhe-2 LvLiang Solution** - Tianhe-2 LvLiang, Intel Xeon E5-2692v2 12C 2.2GHz,	174,720	2,071.4	3,074.5	997

Tulip Trad.: 17
IBM: 45
Alibaba: 66
AWE: 88

Top500.org – July???

6.7

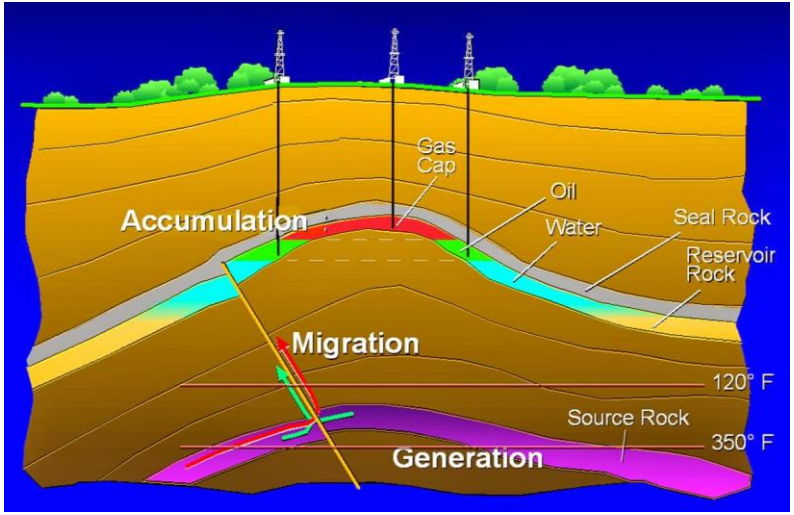
6.0

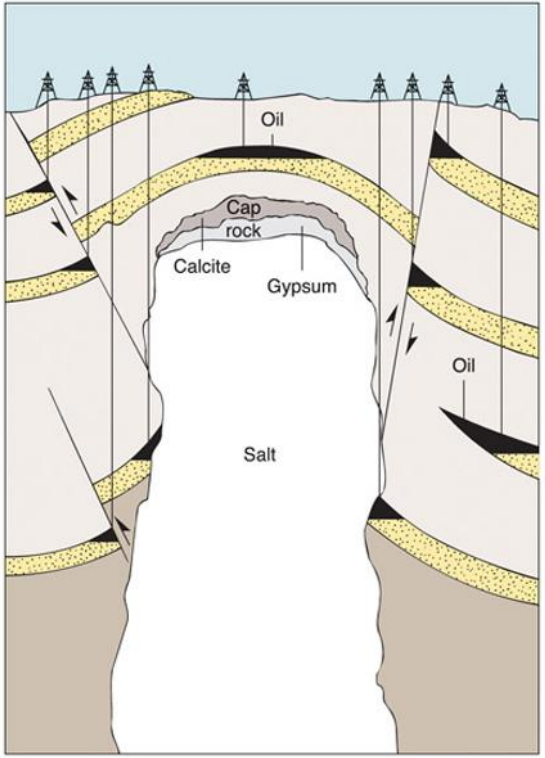
5.0

3.0

2.2

Where do we find hydrocarbons?

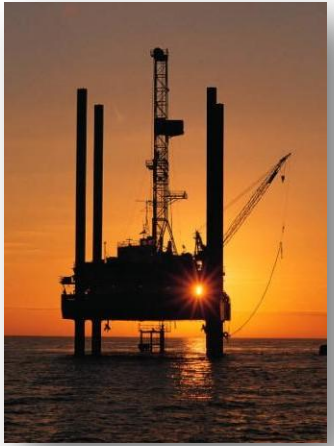




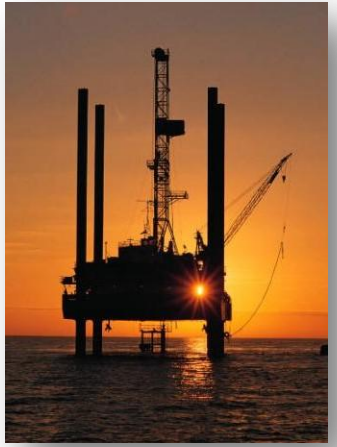
We need **generation**, **migration** and a **trap**

Mostly, "new" oil lies at geologically complex areas

Oil and Gas Exploration



Oil and Gas Exploration



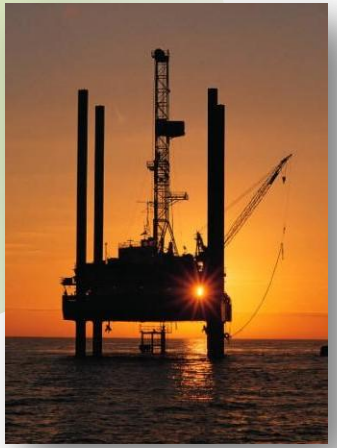
- Land: 1-15 M$
- Shallow water: 30 M$
- Deep water: 100 M$

Oil and Gas Exploration

SUPERCOMPUTING

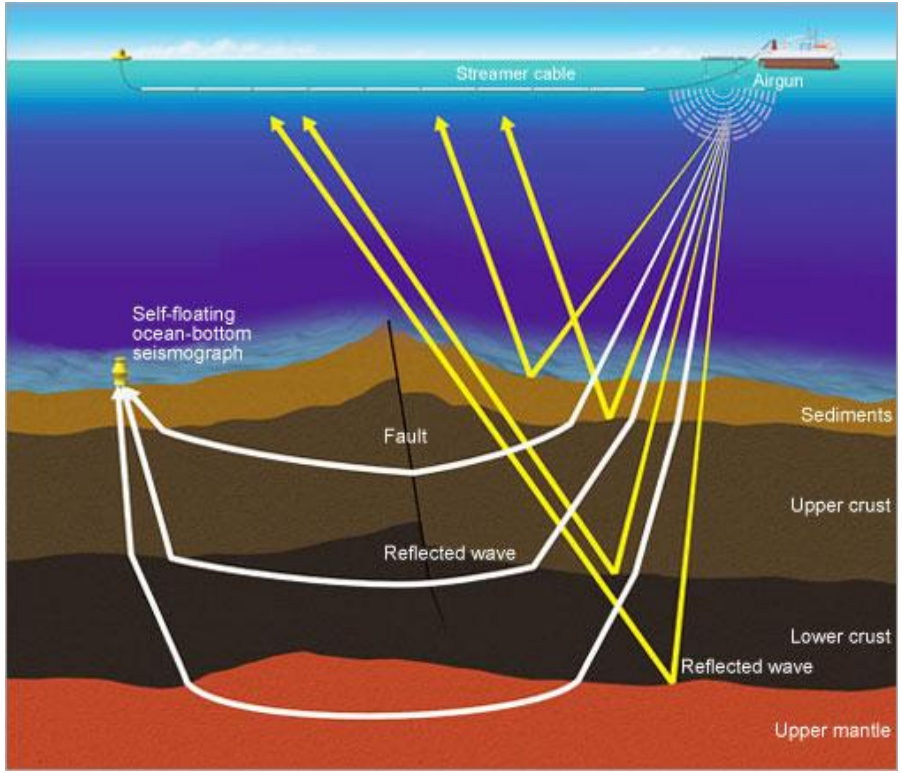
Processing

Interpretation

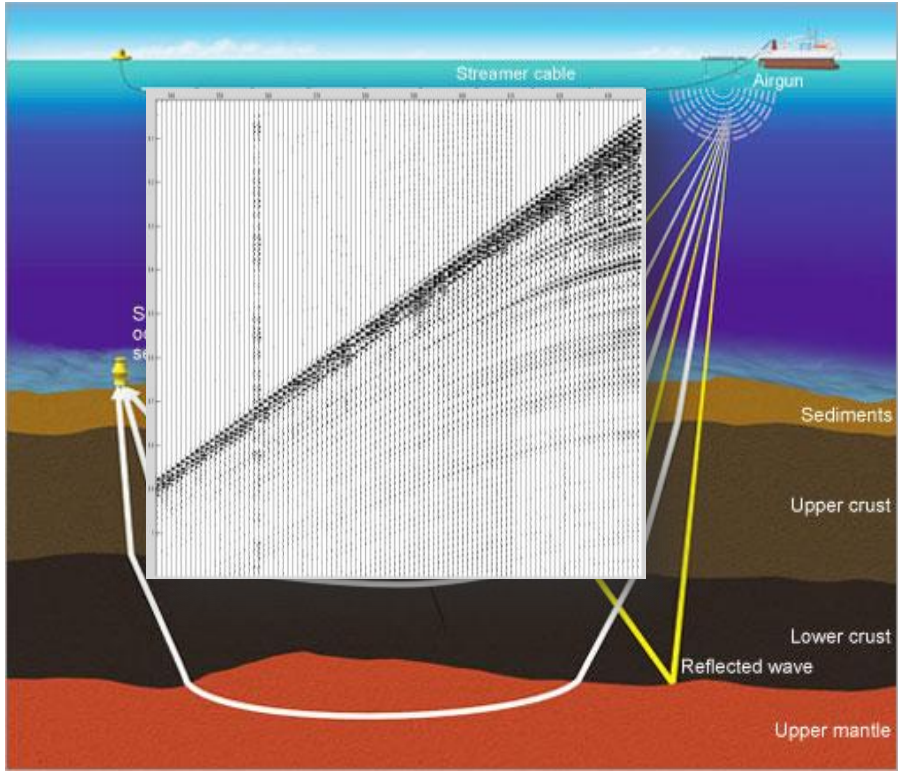


- Land: 1-15 M$
- Shallow water: 30 M$
- Deep water: 100 M$

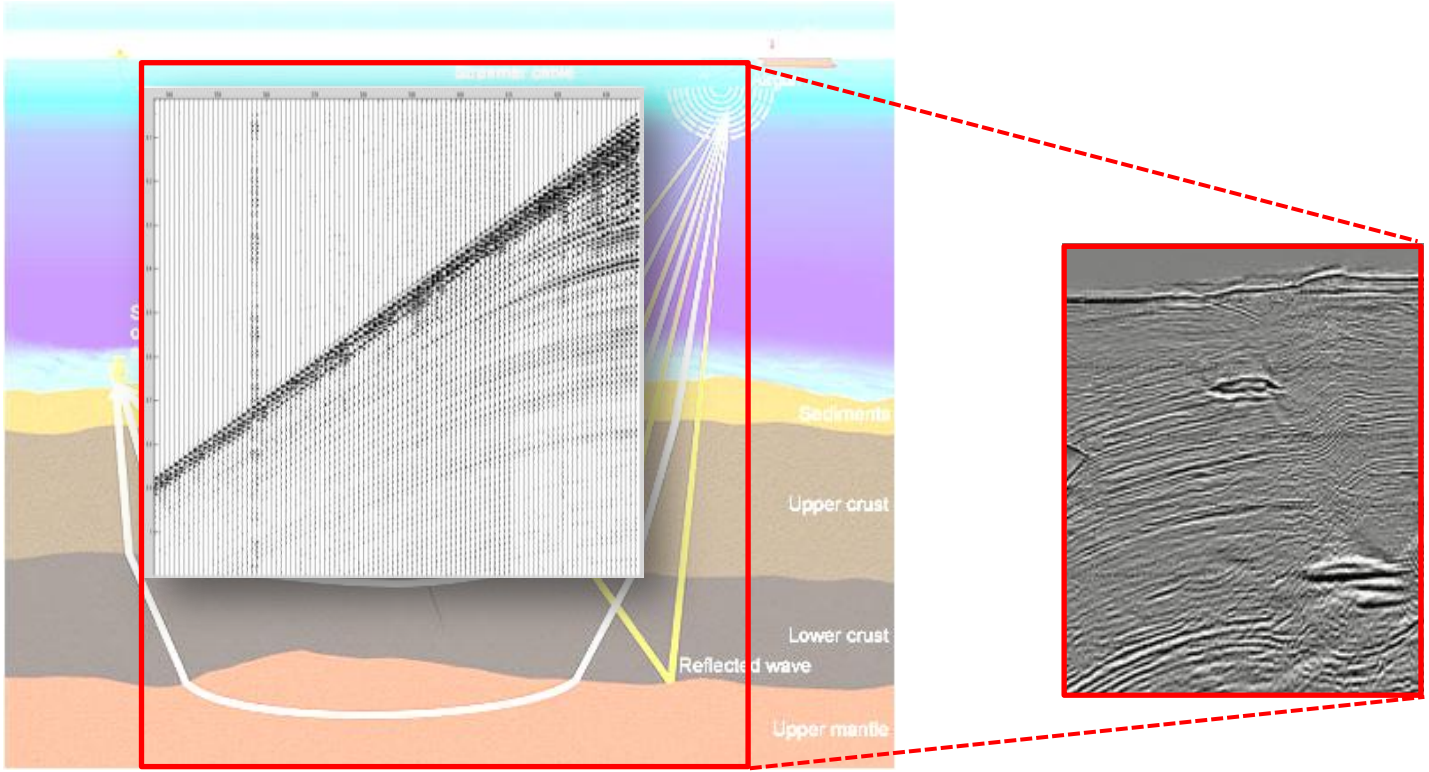
Transforming data into images



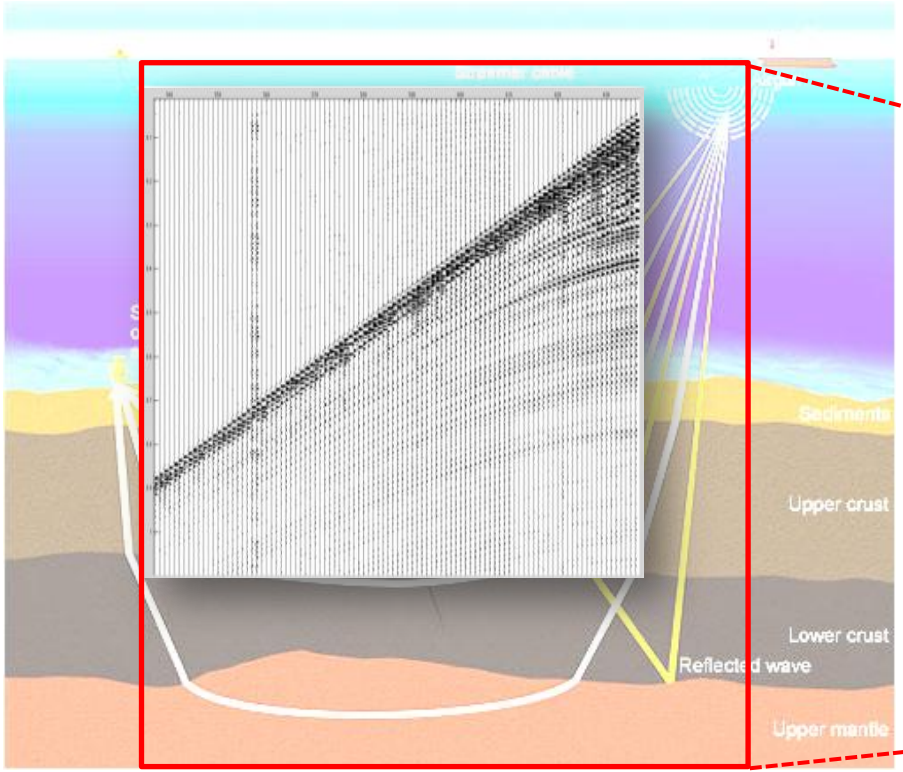
Transforming data into images

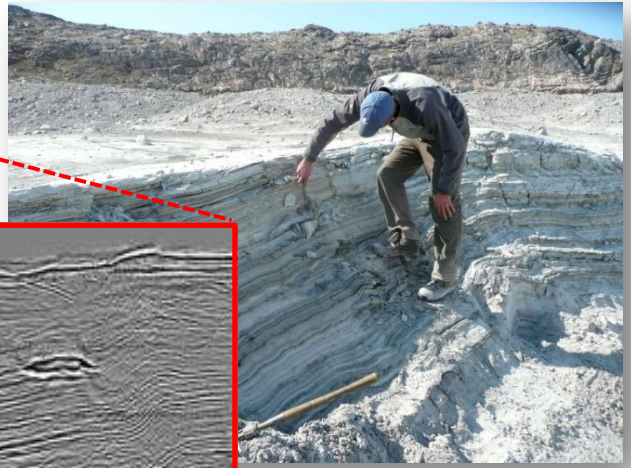


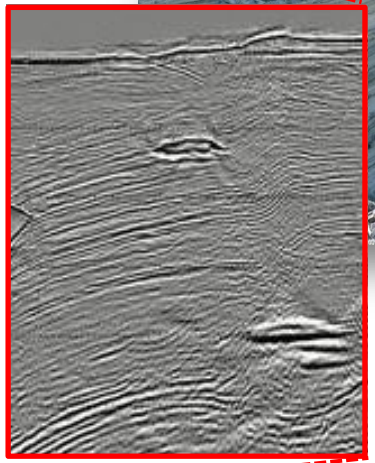
Transforming data into images



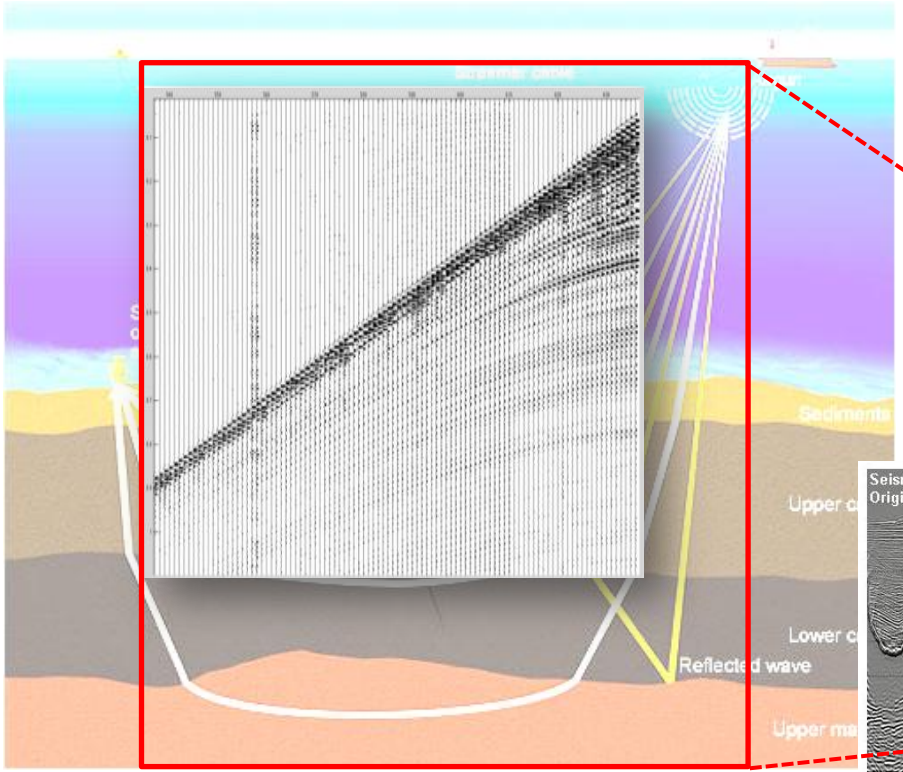
Transforming data into images

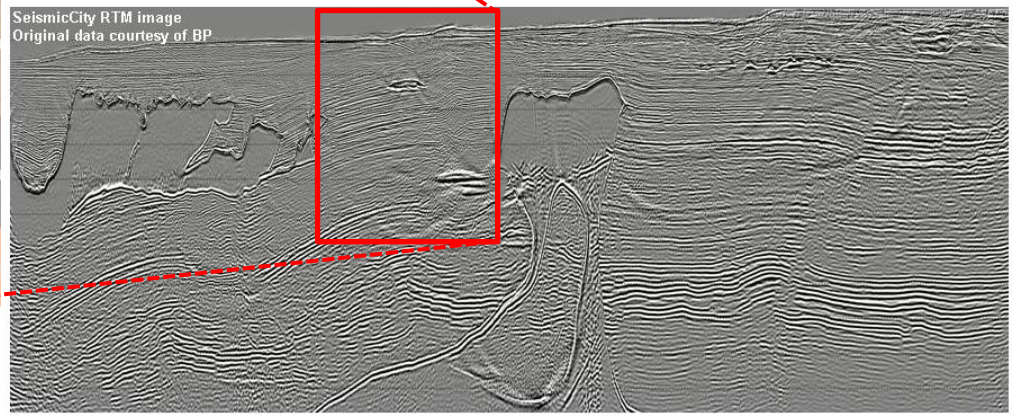




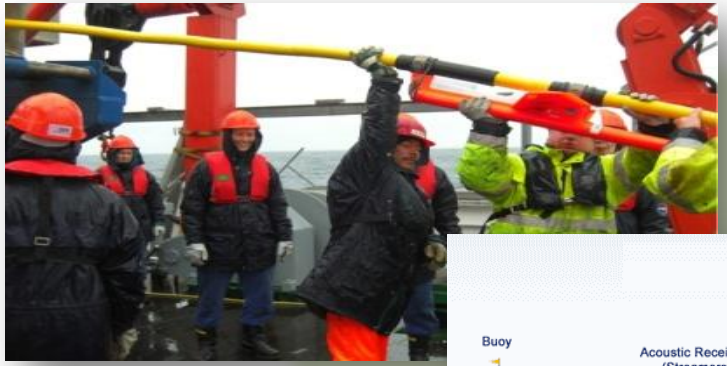


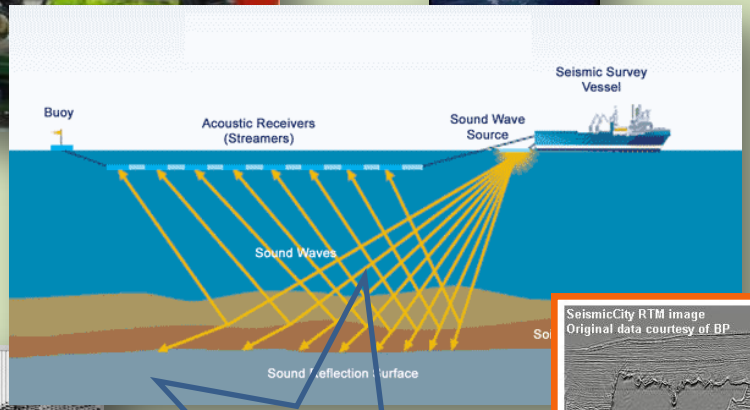
Transforming data into images

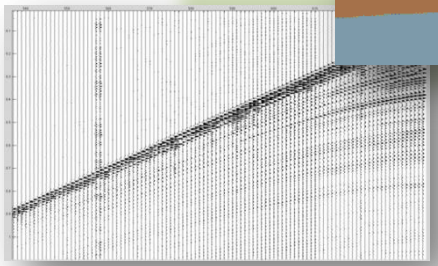




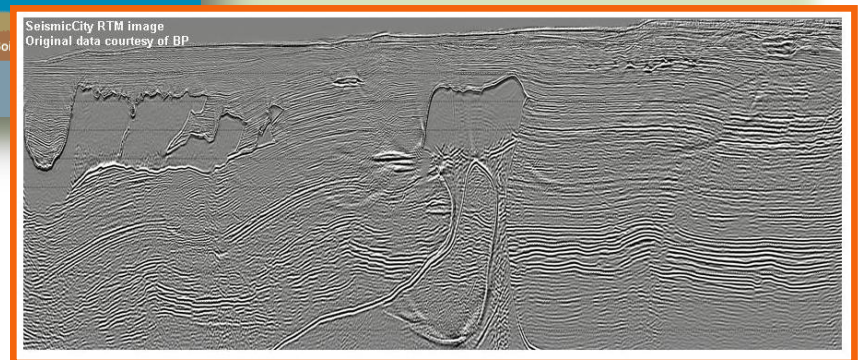
Geophysical Imaging Today







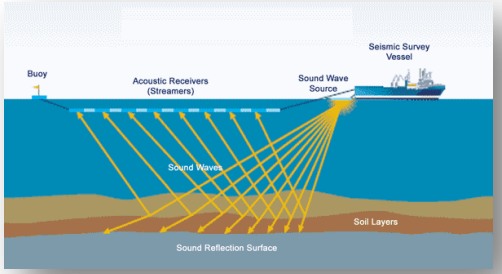
HPC

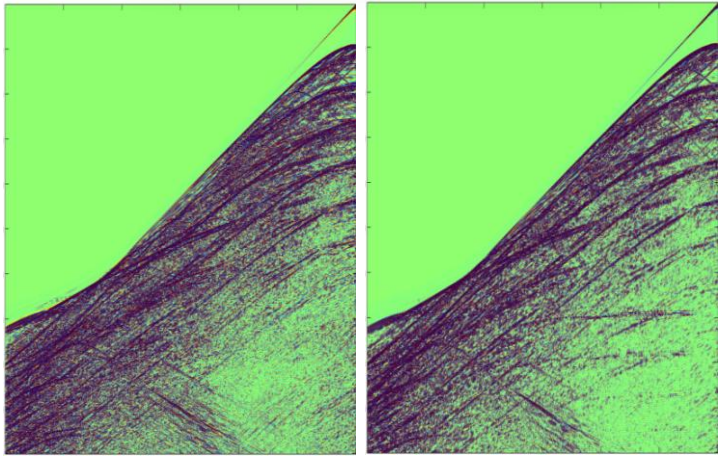



BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación
Forward
FM
Modelling

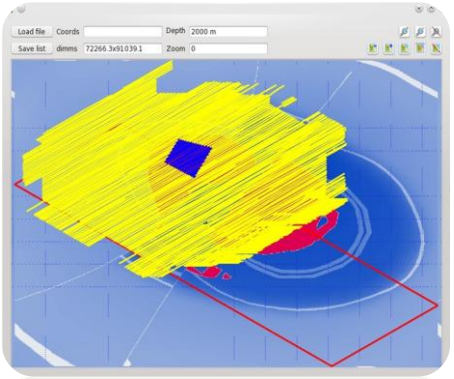
Why Modelling in Hydrocarbon Exploration?

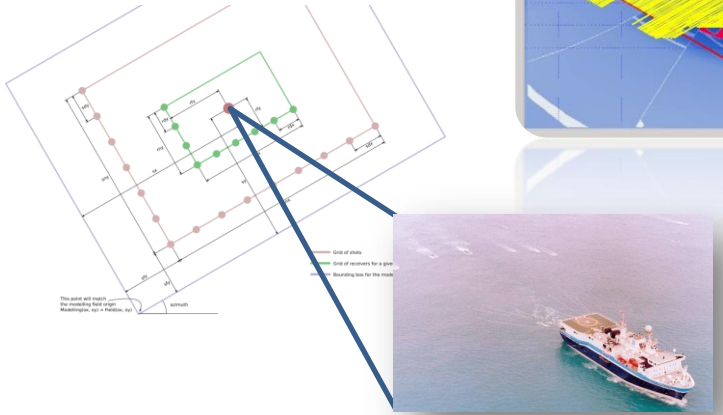
1 - To validate our knowledge of the subsurface



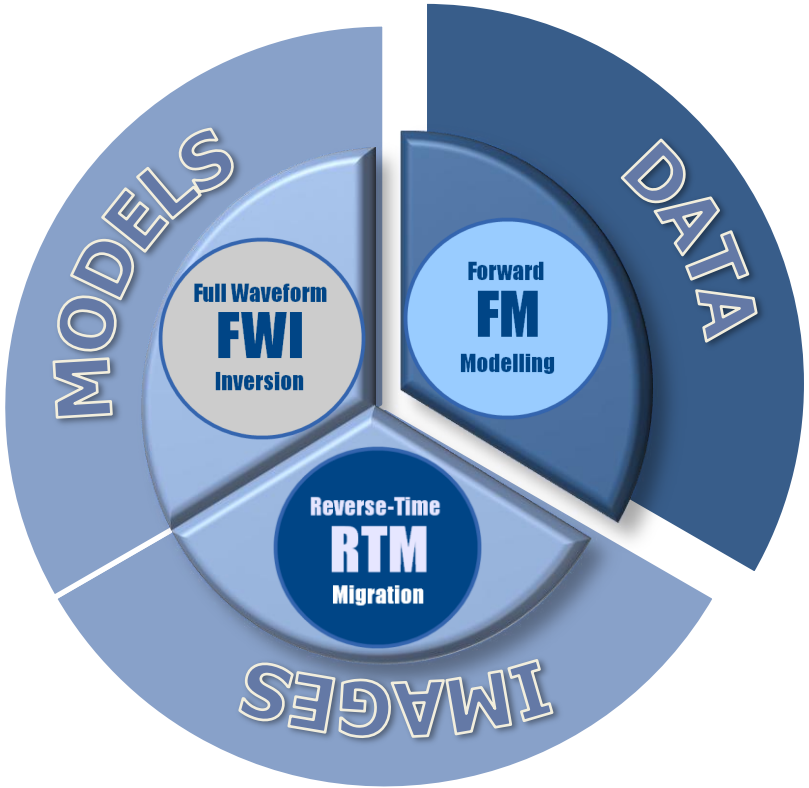


2 - To plan acquisition surveys



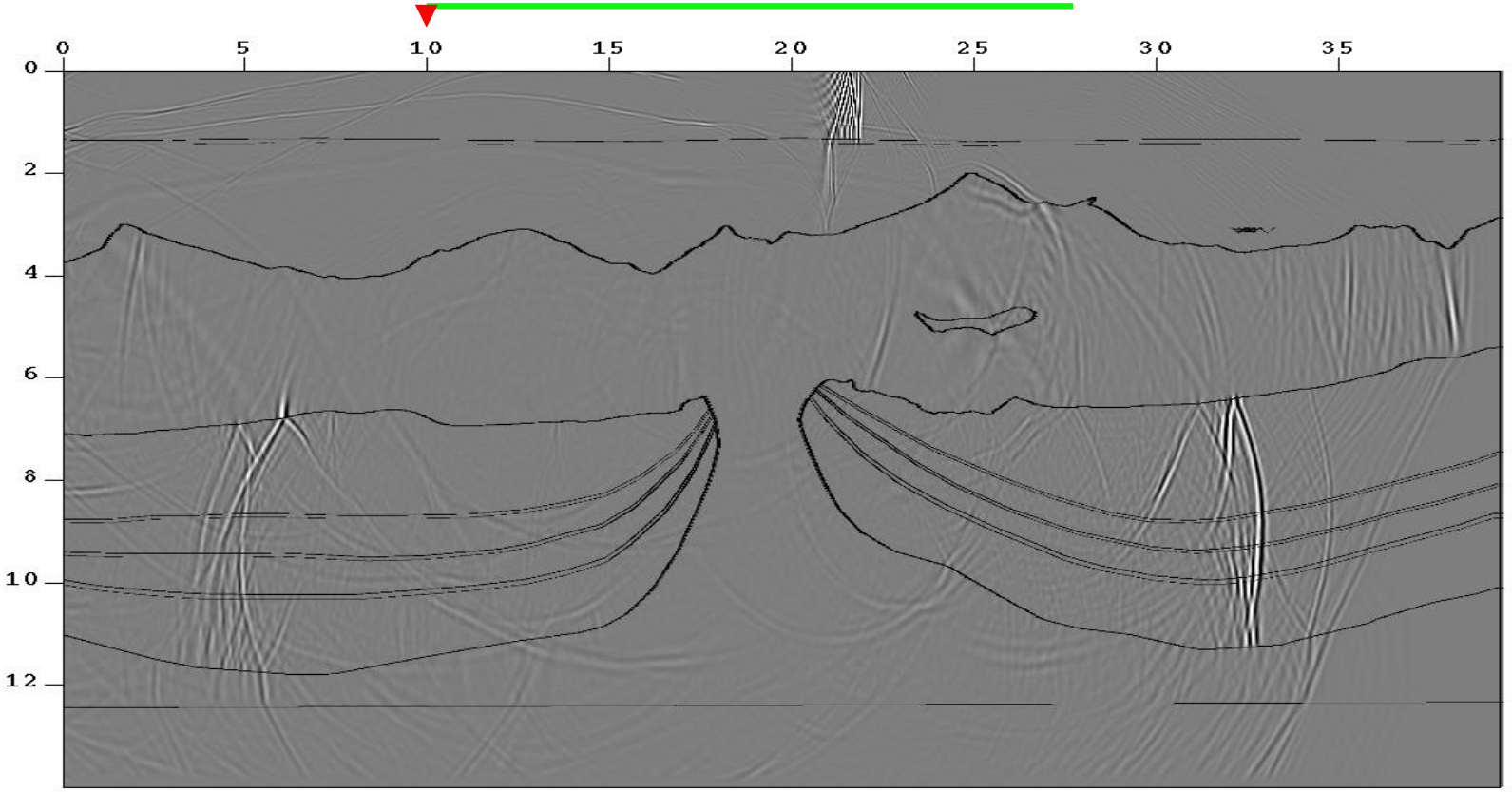


Why Modelling in Hydrocarbon Exploration?



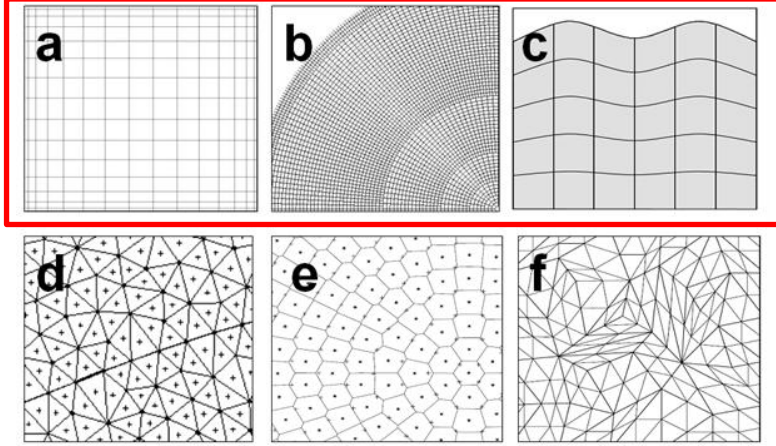
3 - To feed the most advanced 3D imaging engines!

Modelling Complex Wavefields

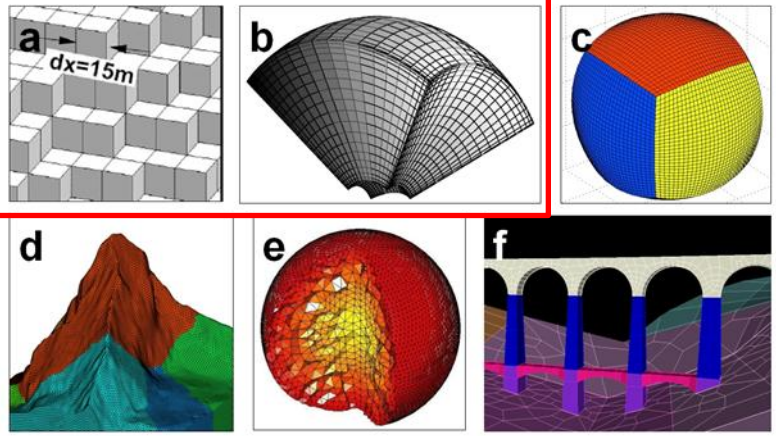


Numerical Methods in the Industry

2D

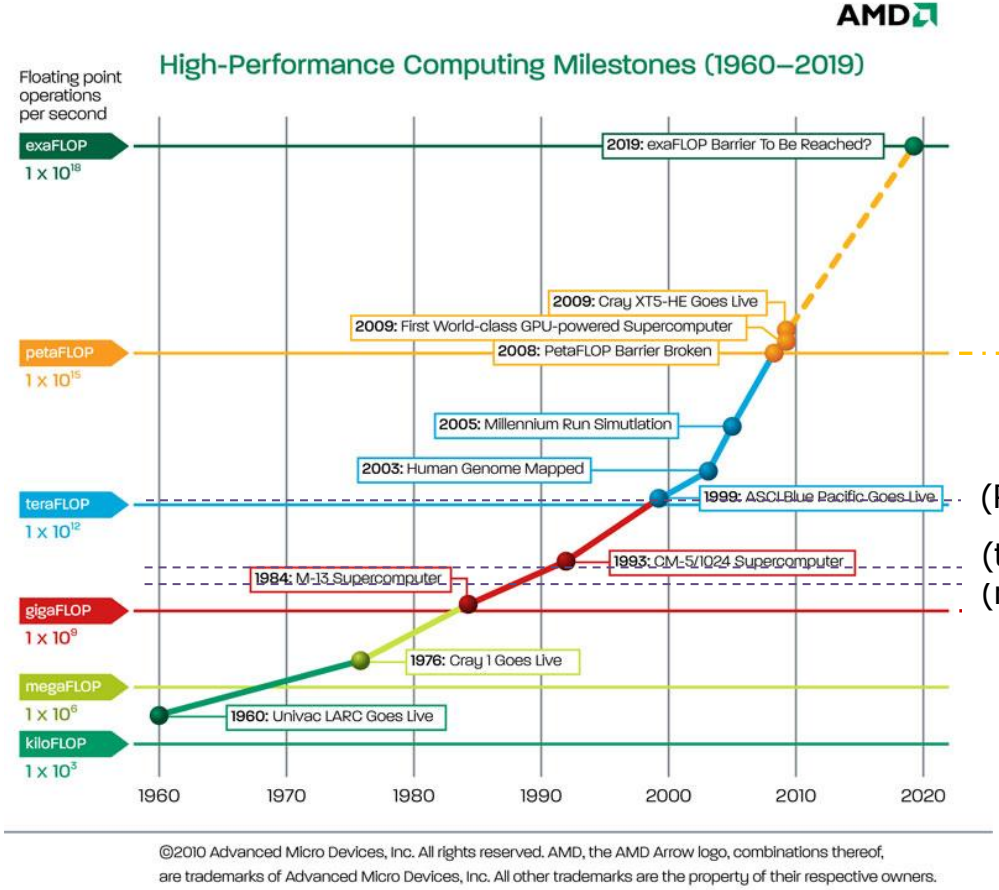


3D:

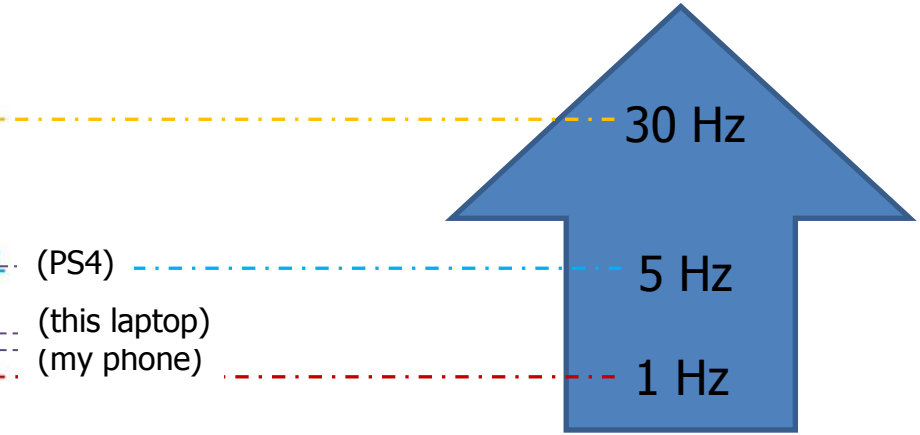


Mostly High-order (spectral) explicit FDTD, with some FDFD
Some Spectral Elements
Few Discontinuous Galerkin

Modelling Cost and Scaling



CPU cost: (frequency) ^{4}
Memory: (frequency) ^{3}



Modelling Cost and Scaling

CPU cost: (frequency) ^{4}
Memory: (frequency) ^{3}

Ratio in Flops of fastest supercomputer w.r.t. fastest commodity? **x1000**

Jump in frequency with a x1000 Flops gain? **x5**

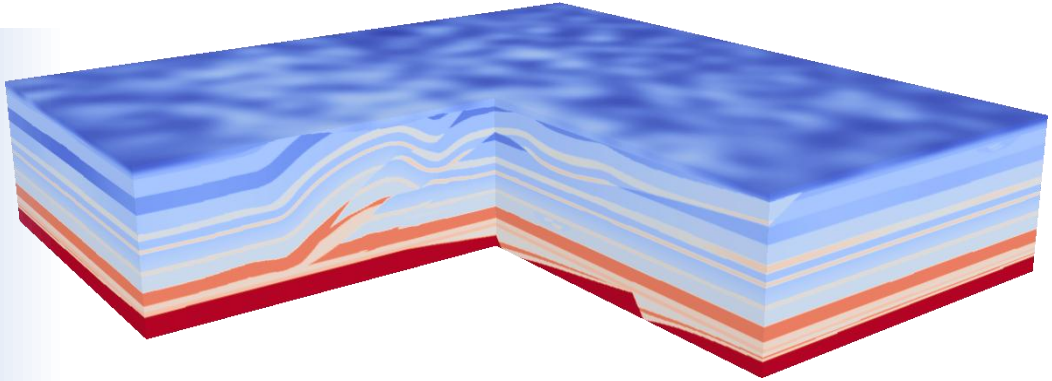
Time until I can perform a x5 jump in frequency with the same computer type? **10 yrs**

Example modelling: SEG/EAGE 3D

- 16 km x 16 km x 3.2 km
- 6 seconds
- 10 Hz peak frequency (25 Hz max.)
- 5041 shots

- 3D linear elastic isotropic model
- FD staggered-grid, explicit time-domain
- Sponge absorbing boundary condition
- Mimetic free surface

- 2.8 TB of output traces

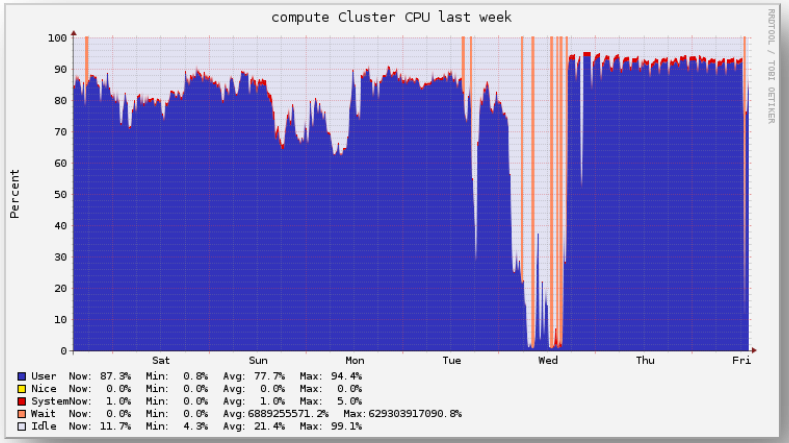


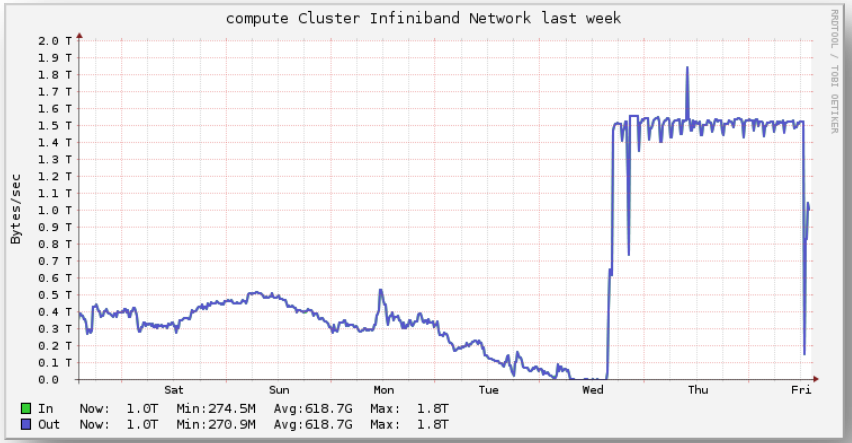
Each shot (kernel):

- 440x2201x2201 cells, 11279 iterations
- 40 cell-long ABCs. Total 121 GB without DD
- ~2.2 hours in 20 nodes
- 6241 seismograms, 4 channels, 39 GB output

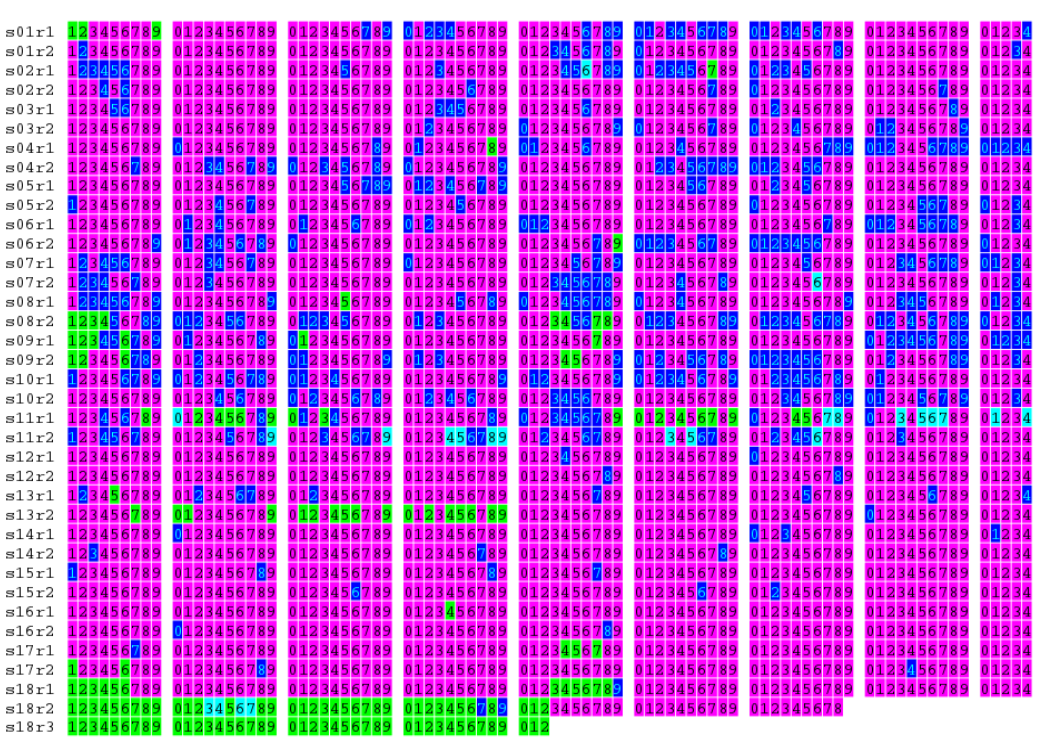
Petascale Seismic Modelling

Using 96% of MareNostrum for 2 days

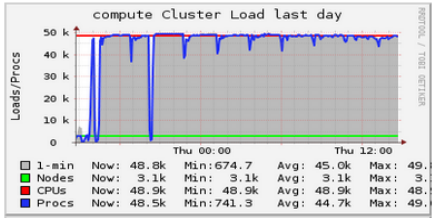


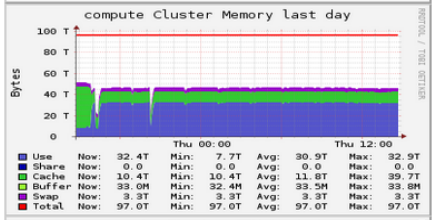


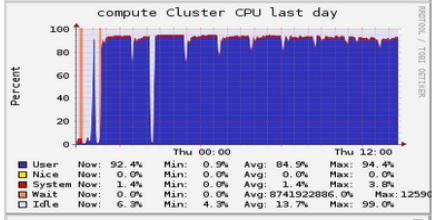
Petascale Seismic Modelling

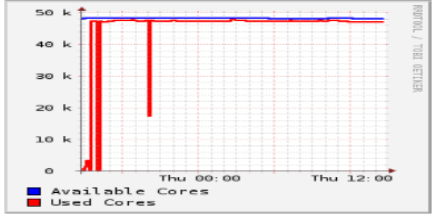


Running Jobs: 28
Waiting Jobs: 440
Blocked Jobs: 2
Dependent Jobs: 0



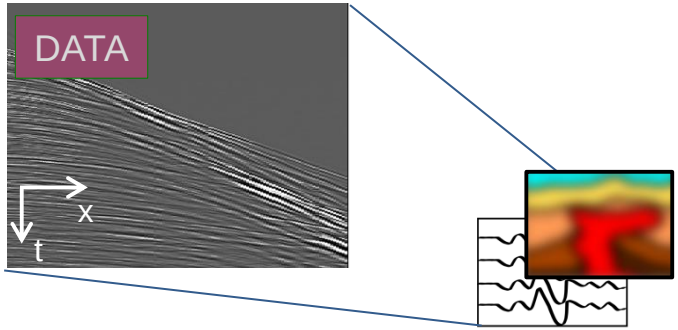


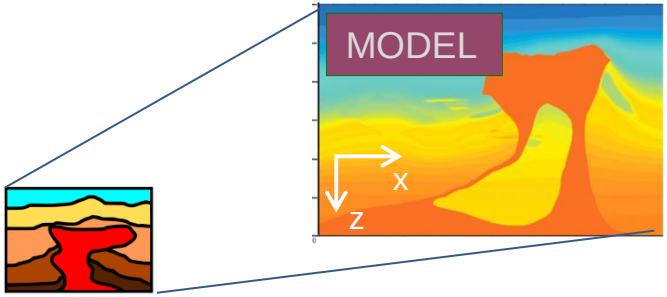


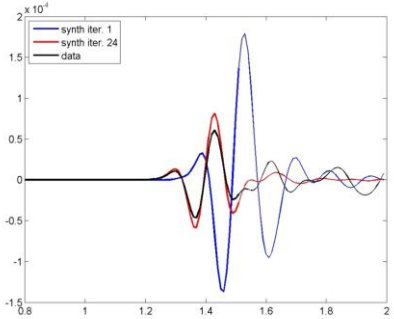



BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación
FULL WAVEFORM INVERSION

FWI: Automatic Model Building

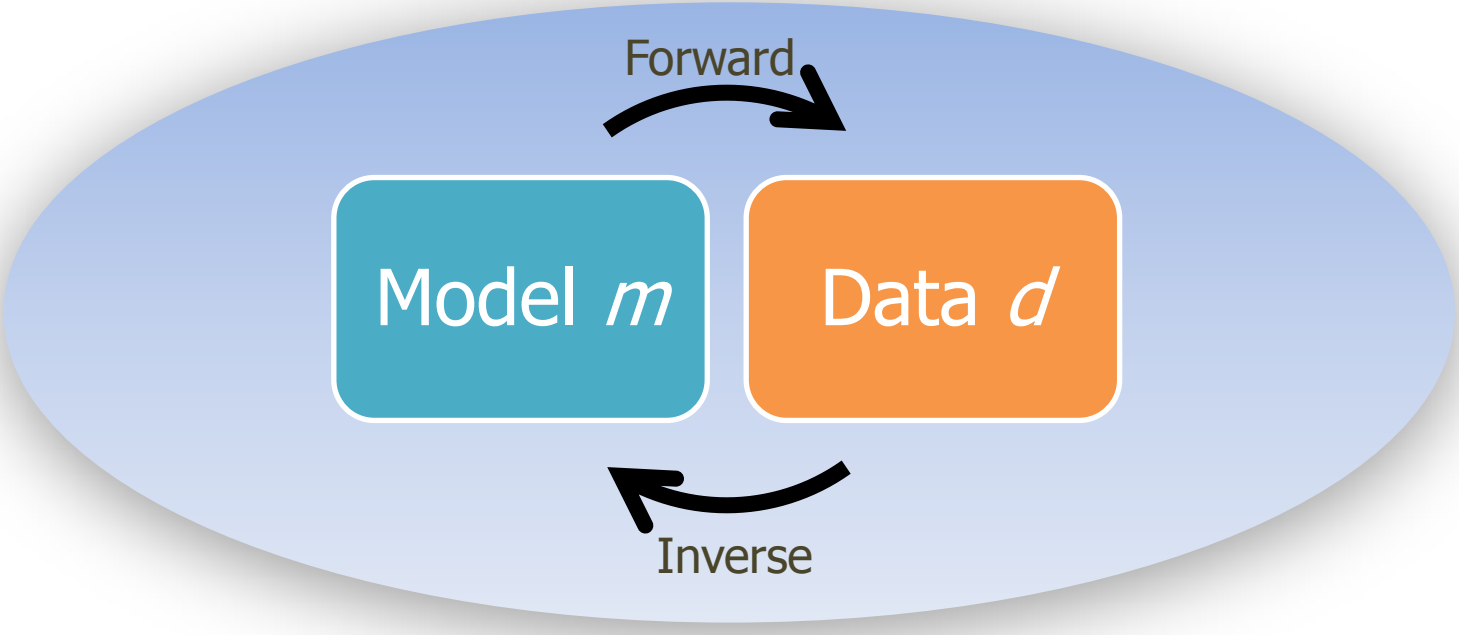






- Unconstrained 3D inversion
- All data can be inverted
- Wave-physics based

Inverse Problem

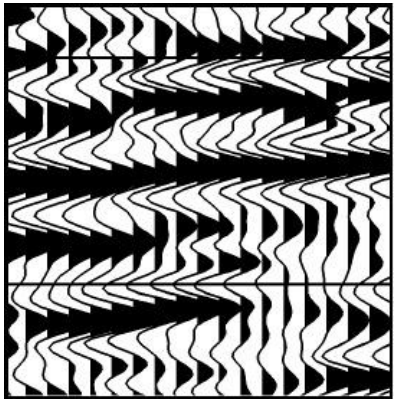


Seismic Inverse Problem

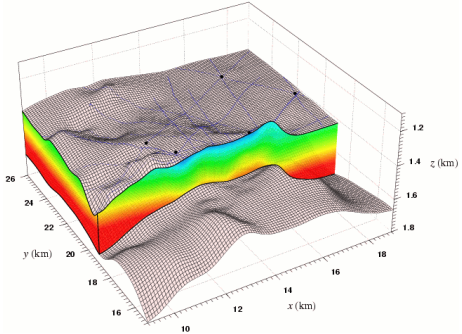
Full Waveform FWI Inversion

BSIT

Seismic data are **seismograms**
(can be pressure, 3C, 4C, …)



Seismic models are **velocities**
(could be density, anisotropic parameters, …)



Adjoint Method

Misfit

$$E = \frac{1}{2}\sum_s \sum_r \int dt[P_{\text{obs}}(\mathbf{x}_r, t; x_s) - P_{\text{cal}}(\mathbf{x}_r, t; x_s)]^2,$$

Gradient

Adjoint wavefield, backpropagated

$$\gamma_n = -\frac{1}{V_n^3}\sum_s \int dt(\partial_t P_f)(\partial_t P_b),$$

Model update

$$V_{n+1} = V_n + \alpha_n \gamma_n$$

Iterative Method (NL-CG)

Steepest slope

$$\Delta m_n = -\nabla f(m_n)$$

Search direction

$$\Lambda m_n = \Delta m_n + \beta_n \Lambda m_{n-1}$$

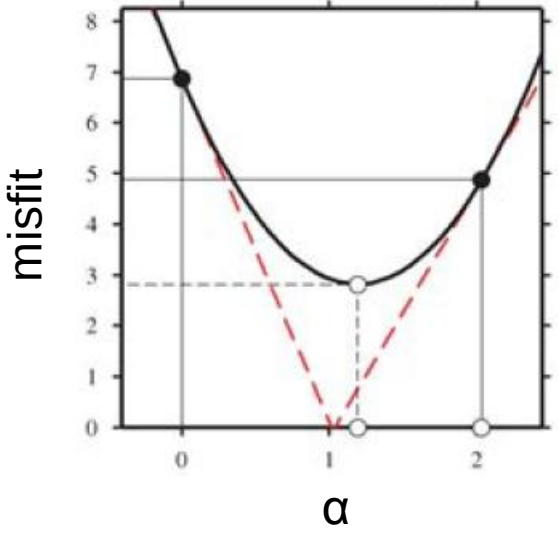
New model

$$m_{n+1} = m_n + \alpha_n \Lambda m_n$$

Search “memory”

$$\beta_n = \beta_n(\Delta m_{n-1}, \Delta m_n)$$

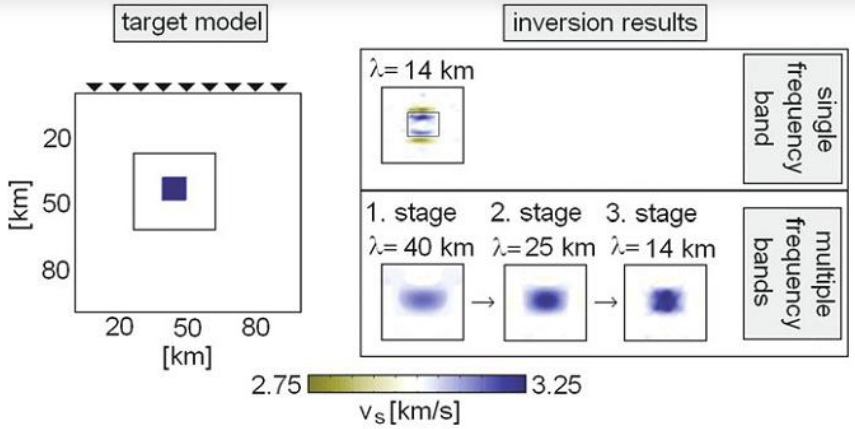
Step size



Barrier 1: Algorithmics

Linear → Non-Linear

Low frequencies	High frequencies
Near offsets	Far offsets
Small time windows	Full time



Barrier 2: Cost

Data: source model, velocity model, receiver traces
Result: Inverted data

1 read and preprocess input data;
2 **for** $f < f_{max}$ **do**
3 **for** $n < n_{max}$ **do**
  `// Compute Gradient (with similar phases and complexity of an RTM)`
4 Compute forward: obtain $\chi(n)$;
5 Compute backward: obtain $g_{model}(k_{model})$;
  `// Minimize misfit (k)`
6 **for** $m < m_{max} || \alpha$ *found* **do**
7 Compute Forward: obtain α ;
8 **end**
  `// Update Model`
9 $\text{model}^{n+1} = f(\text{model}^n, \chi(k), g_{model}(k_{model}), \alpha)$;
10 **end**
11 **end**
12 write inverted variable to disk;

~**RTM**

~**FM**

$\sim n_f \times n \times (\textbf{RTM} + m \times \textbf{FM})$

$n_f \sim 5$
$n \sim 30$ *!!!!*
$m \sim 4$

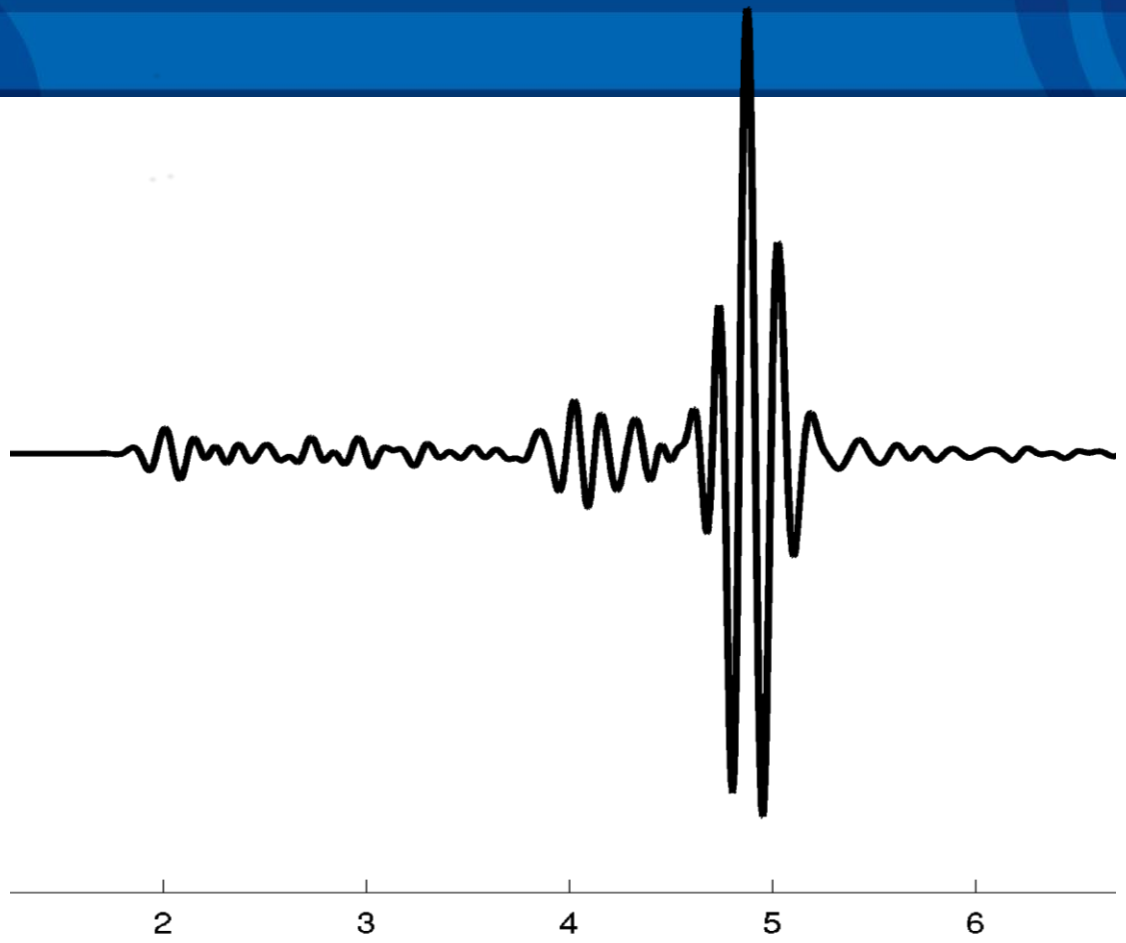
a MASSIVE problem!

Initial guess of A

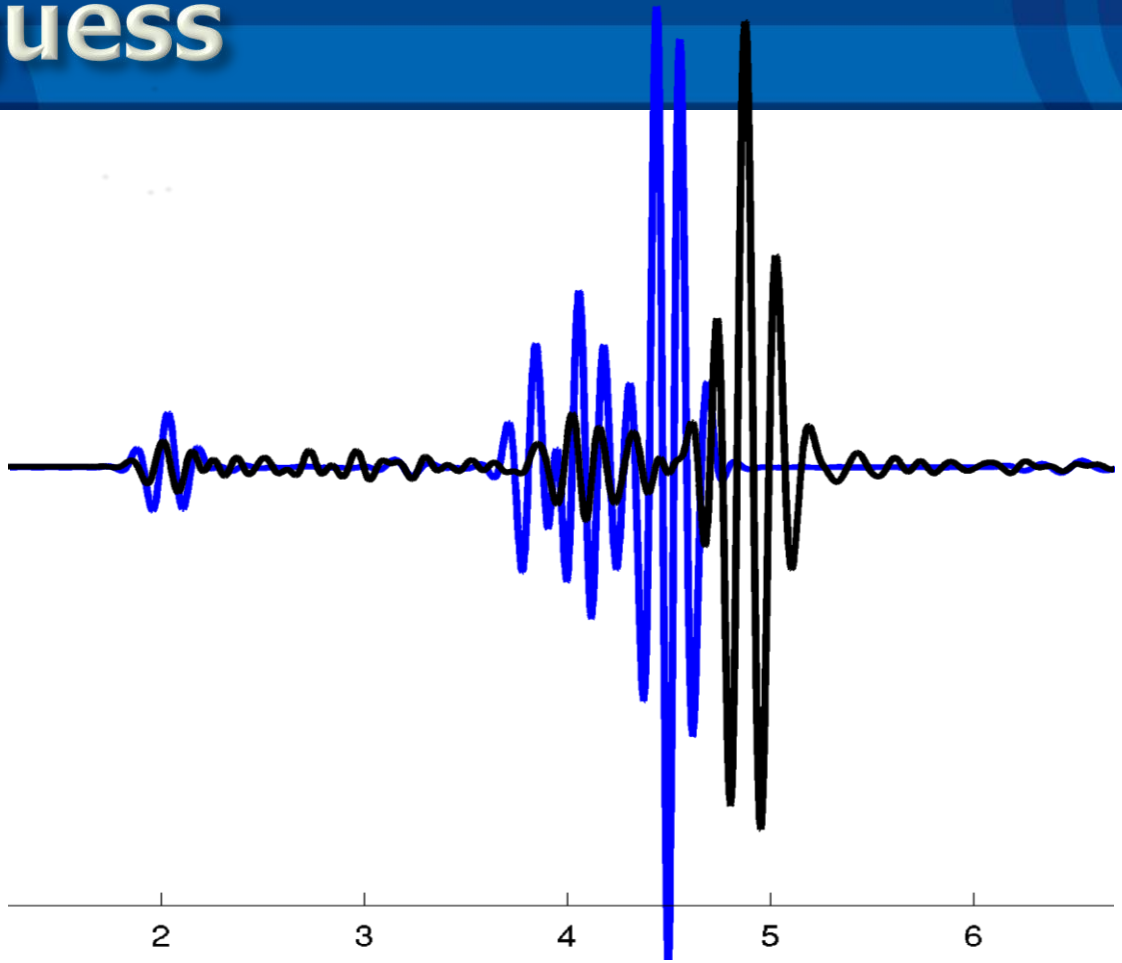
Solve ill-posed, non-linear problem

- Parameters: 160x800x800x2 (≈2e8!)
- Data: 5041x6241x6000x3 (≈5e11!)

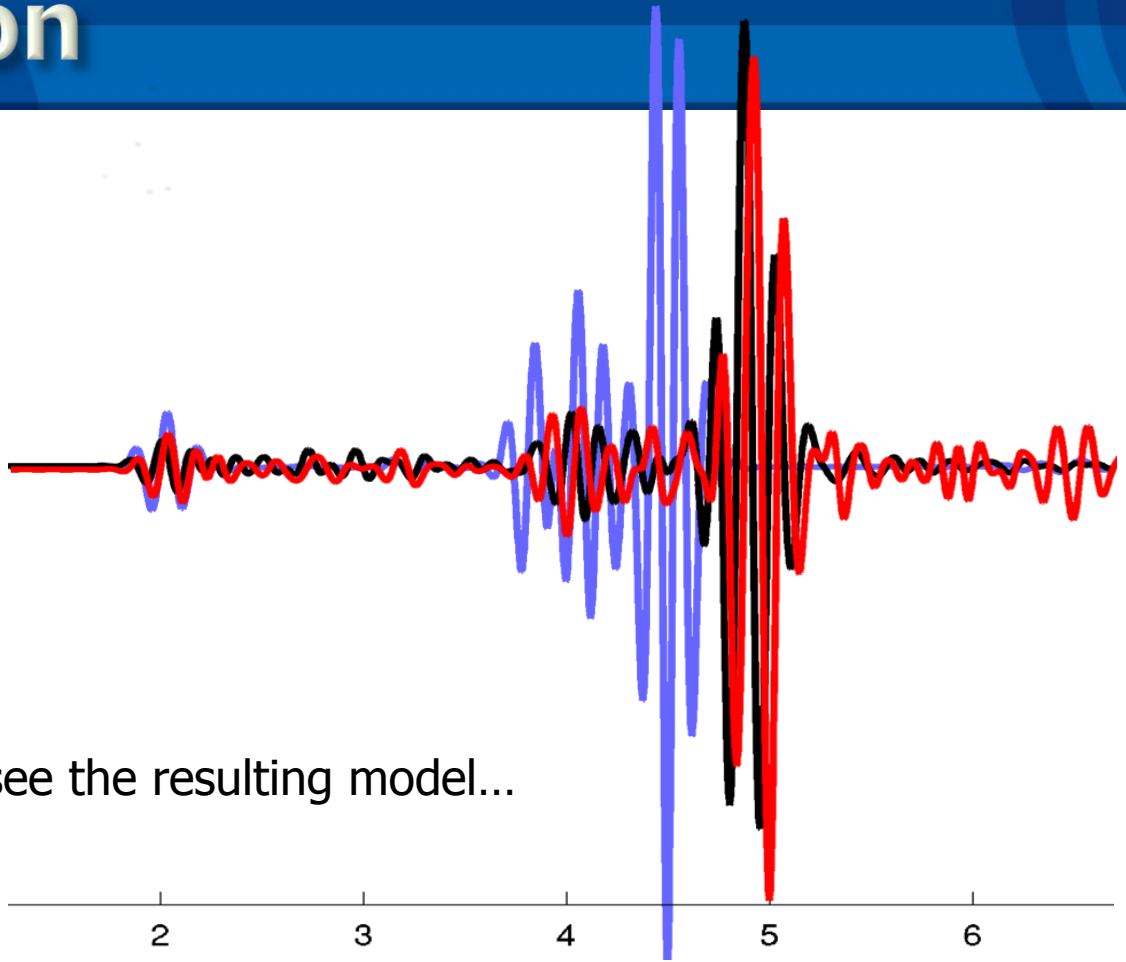
Data



Initial guess



Inversion



…stay tuned to see the resulting model…

CONCLUSION AND FUTURE

Conclusions (part I)

1 – Geophysical information comes from 3D seismic (and EM) surveys.

2 – Modelling (FM) allows us to design surveys and test hypothesis

3 – Inversion (FWI) allows us to retrieve physical information from the subsurface.

4 – Current challenges of the industry involve using the full elastic information in highly complex environments (subsalt, topography, …)

Conclusions (part II)

1 - Geophysics + HPC are the main tool to reduce uncertainty when finding hydrocarbons

2 - In particular, processing seismic data is a grand challenge for computing

3 - The problem to be solved, the physical approximation used and the desired resolution bound the computing resources needed.

OR

3 - The computer resources at hand tell us which approximations to use and which resolutions we will have.

Future Trends

ALGORITHMICS

FM: Unstructured Meshes
FWI: Elastic, Anisotropic, Viscoelastic

WORKFLOWS

Joint EM/Seismic inversion
Near-real time processing
Ultra-high density surveys

COMPUTING

Frequencies: Towards 50 Hz
Shot number: Beyond 100,000
4D simulations
Exascale computing

Visit BSIT!

www.bsc.es/bsit

bsit@bsc.es